

# 基于 ARIMA 和小波神经网络组合模型的交通流预测

成 云,成孝刚,谈苗苗,周 凯,李海波  
(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

**摘 要:**针对现阶段城市道路交通流预测精度不高的局限性,提出了一种基于差分自回归滑动平均 (ARIMA) 和小波神经网络 (WNN) 组合模型的预测方法来进行交通流预测。利用差分自回归滑动平均模型良好的线性拟合能力和小波神经网络模型强大的非线性关系映射能力,把交通流时间序列的数据结构分解为线性自相关结构和非线性结构两部分。采用差分自回归滑动平均模型预测交通流序列的线性部分,用小波神经网络模型预测其非线性残差部分,最终合成为整个交通流序列的预测结果。计算机仿真结果表明:组合模型的预测精度高于 ARIMA 模型和 WNN 模型各自单独使用时的预测精度,组合模型可以提高交通流预测精度,是交通流预测的有效方法。

**关键词:**交通流预测;差分自回归滑动平均模型;小波神经网络;组合模型

中图分类号:U491112

文献标识码:A

文章编号:1673-629X(2017)01-0169-04

doi:10.3969/j.issn.1673-629X.2017.01.038

## Traffic Flow Prediction Based on Hybrid Model of ARIMA and WNN

CHENG Yun, CHENG Xiao-gang, TAN Miao-miao, ZHOU Kai, LI Hai-bo

(School of Telecommunications and Information Engineering, Nanjing University of  
Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** Aimed at the limitation of low prediction accuracy at the present stage of city road traffic, a prediction method is proposed based on Hybrid Autoregressive Integrated Moving Average (ARIMA) and Wavelet Neural Network (WNN) to predict traffic flow. Using the good linear fitting ability of ARIMA and the strong nonlinear mapping ability of WNN, the traffic flow time series are considered to be composed of a linear autocorrelation structure and a nonlinear structure. ARIMA model is used to predict the linear component of traffic flow time series and the wavelet neural network model is applied to the nonlinear residual component prediction. The simulation results show that the hybrid model can produce more accurate prediction than that of single model, which improves prediction accuracy of traffic flow prediction, and it's an efficient method.

**Key words:** traffic flow prediction; ARIMA model; wavelet neural network; hybrid model

## 0 引 言

准确实时的交通流信息在智能交通系统中起着十分重要的作用,因此交通流预测受到广泛关注<sup>[1]</sup>。其主要预测方法包括 3 种:基于线性理论的预测方法、基于知识发现的智能模型预测方法、基于非线性理论的预测方法<sup>[2]</sup>。其中一些预测方法在交通流预测应用中取得了不错的预测效果,但是单个预测方法并不适用于所有交通状态,同时实际的交通流往往既有线性特征,又有非线性特征,仅采用某一种预测模型,难以很

好地反映出时间序列的线性与非线性双重特征,所以应该依据交通流的实际情况来选取合适的预测模型。为了更好地发挥各种模型的预测优势,在 1969 年 Bates 等提出组合预测理论,即把多种预测方法组合起来以获得更好的预测效果。Zhang 在预测太阳黑子的问题中也采用了组合模型,其预测效果比单个模型更好<sup>[3]</sup>。

文中使用差分自回归滑动平均 (ARIMA) 模型和小波神经网络 (WNN) 模型相结合的方法来预测交通

收稿日期:2016-03-27

修回日期:2016-06-28

网络出版时间:2017-01-04

基金项目:国家自然科学基金资助项目(61401236);南京邮电大学引进人才项目(NY214005)

作者简介:成 云(1994-),女,研究方向为交通流预测在智能交通系统中的应用;成孝刚,讲师,研究方向为交通流预测在智能交通系统中的应用和雾霾能见度检测。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170104.1039.072.html>

流,利用两种模型各自的优点来提高交通流的预测精度。

## 1 交通流预测模型

### 1.1 ARIMA 模型

#### 1.1.1 ARIMA 基本理论

ARIMA 是由自回归滑动平均模型 (ARMA) 扩展而来。该模型首先对非平稳时间序列作多次差分使其成为平稳时间序列,  $d$  即为差分次数,再对差分后的序列建立参数为  $p$ 、 $q$  的 ARMA 模型,然后由反变换得到原来的非平稳时间序列<sup>[4-5]</sup>。参数为  $p$ 、 $d$ 、 $q$  的 ARIMA 模型预测方程如下:

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

其中,  $y_t$  是样本值;  $\varphi_i (i=1, 2, \cdots, p)$  和  $\theta_i (i=1, 2, \cdots, q)$  是模型的参数;  $\varepsilon_t$  是白噪声序列,其服从独立正态分布  $N(0, \sigma_a^2)$ 。

$p$ 、 $d$ 、 $q$  是模型的阶数,由不同的  $p$ 、 $d$ 、 $q$  组合测试可以找到最合适的模型参数<sup>[6]</sup>。

#### 1.1.2 ARIMA 算法步骤

ARIMA 建模与预测包含四个步骤<sup>[7]</sup>,如下所示:

(1) 序列平稳化处理。

(2) 模型识别。一般根据自相关系数与偏自相关系数来确定 ARIMA 模型的阶数  $p$  和  $q$ 。

(3) 参数估计和模型诊断。

(4) 用所选取参数的模型来预测时间序列。

### 1.2 小波神经网络模型

#### 1.2.1 小波神经网络基本理论

WNN 是以误差反传神经网络拓扑结构为基础的网络,与 BP 神经网络类似。在小波神经网络中,信号向前传播的同时,误差反向传播,不同的是小波神经网络隐含层节点的传递函数为小波基函数<sup>[8]</sup>。WNN 拓扑结构如图 1 所示。

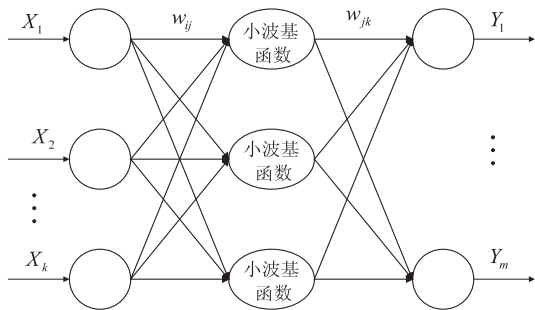


图 1 WNN 拓扑结构

图 1 中,  $X_1, X_2, \cdots, X_k$  是 WNN 的输入参数;  $Y_1, Y_2, \cdots, Y_m$  是预测的输出值,即需要求得的最终结果;  $w_{ij}$  和  $w_{jk}$  为网络连接权重值。

当输入交通流量样本序列为  $x_i (i=1, 2, \cdots, k)$  时,

隐含层输出为:

$$h(j) = h_j \left( \frac{\sum_{i=1}^k w_{ij} x_i - b_j}{a_j} \right), j=1, 2, \cdots, l \quad (2)$$

其中,  $h(j)$  为隐含层第  $j$  个节点的输出;  $w_{ij}$  为输入层和隐含层的连接权重值;  $a_j$  为小波基函数  $h_j$  的伸缩因子;  $b_j$  为  $h_j$  的平移因子;  $h_j$  为小波基函数<sup>[9]</sup>。

文中的小波基函数是 Morlet 母小波基函数,公式为:

$$y = \cos(1.75x) e^{-\frac{x^2}{2}} \quad (3)$$

小波神经网络输出层计算公式为:

$$y(k) = \sum_{i=1}^l w_{ik} h(i), k=1, 2, \cdots, m \quad (4)$$

其中,  $w_{ik}$  为隐含层到输出层的网络连接权重值;  $h(i)$  为第  $i$  个隐含层节点的输出;  $l$  为隐含层节点个数;  $m$  为输出层节点个数<sup>[10]</sup>。

小波神经网络权值修正通过梯度修正法来修正网络的权值和小波函数参数,使得网络预测输出不断接近期望输出。步骤如下:

(1) 计算网络预测误差。

$$e = \sum_{k=1}^m y_n(k) - y(k) \quad (5)$$

其中,  $y(k)$  为交通流的预测输出结果;  $y_n(k)$  为交通流的期望输出结果。

(2) 通过预测误差  $e$  修正相关系数和网络权重值。

$$w_{n,k}^{(i+1)} = w_{n,k}^{(i)} + \Delta w_{n,k}^{(i+1)} \quad (6)$$

$$a_k^{(i+1)} = a_k^{(i)} + \Delta a_k^{(i+1)} \quad (7)$$

$$b_k^{(i+1)} = b_k^{(i)} + \Delta b_k^{(i+1)} \quad (8)$$

其中, 根据网络预测误差计算得到  $\Delta w_{n,k}^{(i+1)}$ 、 $\Delta a_k^{(i+1)}$ 、 $\Delta b_k^{(i+1)}$ 。

$$\Delta w_{n,k}^{(i+1)} = -\eta \frac{\partial e}{\partial w_{n,k}^{(i)}} \quad (9)$$

$$\Delta a_k^{(i+1)} = -\eta \frac{\partial e}{\partial a_k^{(i)}} \quad (10)$$

$$\Delta b_k^{(i+1)} = -\eta \frac{\partial e}{\partial b_k^{(i)}} \quad (11)$$

其中,  $\eta$  代表学习速率。

#### 1.2.2 小波神经网络算法步骤

WNN 训练算法步骤<sup>[11]</sup>如下:

(1) 网络初始化。需要初始化的有伸缩因子、平移因子、网络权重值、学习速率。

(2) 样本分类。把样本分为训练样本和测试样本。

(3) 预测输出。输入训练样本,获得预测输出,计算预测输出与期望输出的误差。

(4) 进行权值修正。通过误差修正网络权重值和小

波函数参数,使误差控制在一定范围之内。

(5)判断算法是否结束,如果没有,返回步骤(3)。

网络训练完成后,输入测试样本,根据预测输出结果与测试样本输出的对比来评估 WNN 的性能。综上所述,基于 WNN 的交通流预测流程如图 2 所示。

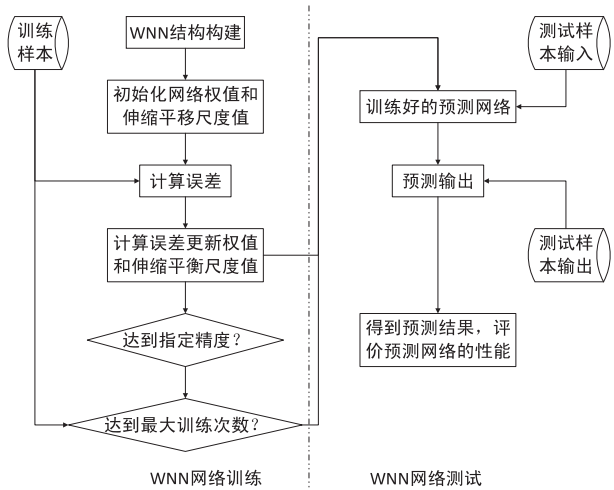


图 2 基于 WNN 的交通流预测流程

1.3 组合模型

实际的时间序列往往既有线性特征,又有非线性特征<sup>[12-13]</sup>,仅用 ARIMA 或 WNN 都不能很好地反映出时间序列的线性与非线性双重特征。因此,它们的组合模型的预测精度可能比其单一模型更高。

把交通流时间序列的数据  $y_t$  分解为线性  $L_t$  和非线性  $N_t$ <sup>[14-15]</sup>,即:

$$y_t = L_t + N_t \tag{12}$$

组合模型预测包含三个步骤:

(1)采用 ARIMA 模型预测,设预测结果为  $\hat{L}_t$ ,原序列  $y_t$  和预测结果  $\hat{L}_t$  的残差为  $e_t$ ,即:

$$e_t = y_t - \hat{L}_t \tag{13}$$

原交通流时间序列中的非线性关系隐含在序列  $\{e_t\}$  中,即:

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \varepsilon_t \tag{14}$$

其中,  $\varepsilon_t$  表示随机误差。

(2)用 WNN 模型预测序列  $\{e_t\}$ ,设 WNN 模型的预测结果为  $\hat{N}_t$ 。

(3)用 ARIMA 和 WNN 模型进行组合预测,设组合预测结果为  $\hat{y}_t$ ,则:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{15}$$

2 计算机仿真结果

实验中的交通流数据是运用美国加州高速公路通行能力度量系统 (Caltrans Performance Measurement System, PeMS) 数据库中连续五天的交通数据,数据采样间隔为 5 min,数据总数量达到 1 440,选择前四天的

数据作为样本,对第五天的交通流进行预测。

方法 1:单独使用 ARIMA 模型进行预测。首先对原始交通流数据作平稳化处理,再根据已平稳的时间序列的自相关函数与偏自相关函数图以及 AIC 最小准则,获得模型参数  $p, d, q$ ,然后利用该模型进行预测。用 Matlab 实现仿真。

方法 2:单独使用 WNN 模型进行预测,文中的 WNN 结构为 3-8-1,即输入层 3 个节点,隐含层 8 个节点,输出层 1 个节点,网络反复训练 500 次。将前三天的数据用于训练网络,第四天的数据用于测试网络。这一过程使用 Matlab 实现。

方法 3:使用组合模型来预测,步骤见第 1.3 节,其中的步骤二与方法 2 中不同,这一步使用的训练样本和测试样本均来自  $\{e_t\}$ ,见式(13)。

使用平均绝对百分误差 (MAPE) 来评估算法的预测性能。

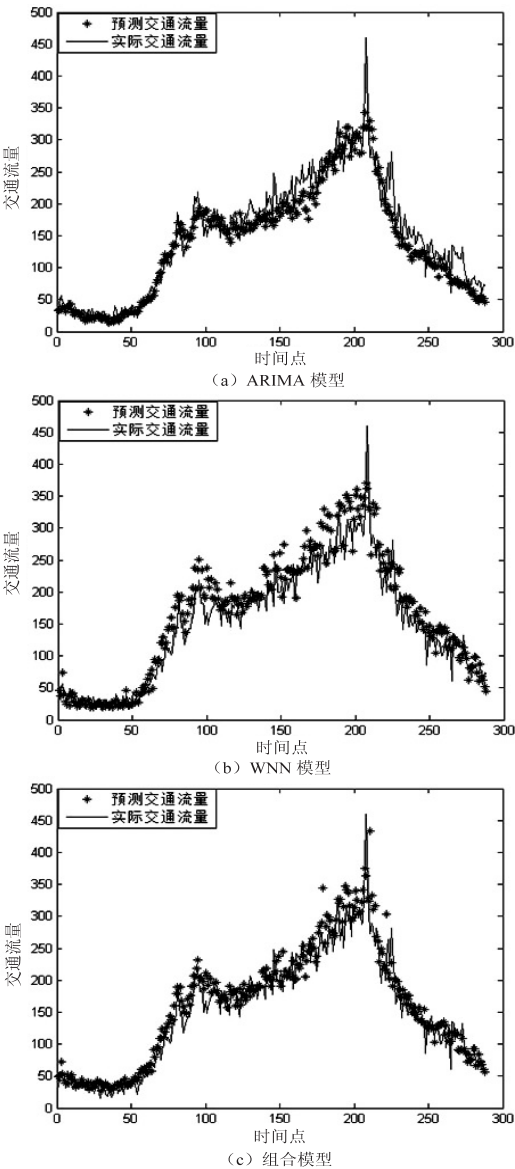


图 3 3 种模型的交通流预测仿真结果

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%$$

(16)

其中,  $n$  为预测点的个数;  $y_i$  为交通流的实际测量值;  $\hat{y}_i$  为交通流的仿真预测值。

MAPE 越小, 预测误差越小, 模型的预测效果就越好。

用上述 3 种方法仿真(ARIMA 模型、WNN 模型、组合模型)获得的交通流预测值和交通流实际测量值的对比见图 3。3 种算法的平均绝对百分误差对比见表 1。

表 1 3 种方法的平均绝对百分误差对比

方法	MAPE/%
ARIMA 模型	4.35
WNN 模型	4.47
组合模型	3.95

由表 1 可得:组合模型的 MAPE 最小,表明组合模型的预测效果最好。

3 结束语

在交通流预测中,由于实际的交通流往往既有线性特征,又有非线性特征,仅采用某一种预测模型,难以很好地反映出时间序列的线性与非线性双重特征。因此提出一种 ARIMA 与 WNN 的组合模型来预测交通流。用 ARIMA 模型预测交通流序列的线性部分,用 WNN 模型预测其非线性残差部分,最后相加得到整个序列的预测结果。通过计算机仿真实验,证明该组合模型比单个预测模型有更好的预测效果。

参考文献:

[1] Daraghmi Y A, Yi C W, Chiang T C. Negative binomial additive models for short-term traffic flow forecasting in urban areas[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(2):784-793.

[2] Teselentis D I, Vlahgianni E I, Karlaftis M G. Improving short-term traffic forecasts: to combine models or not to combine? [J]. IET Intelligent Transport Systems, 2014, 9(2):193-201.

[3] Zhang D C, Wang S F. A novel long term traffic forecast algorithm and case study for China[C]//International conference on advanced information networking and applications workshops. [s. l.]:[s. n.], 2014.

[4] Chan K Y, Dillon T S, Singh J, et al. Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg-Marquardt algorithm [J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(2):644-654.

[5] Wei W, Ma H. ARMA model and wavelet-based ARMA model application [J]. Applied Mechanics and Materials, 2011, 121-126:1799-1803.

[6] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction [J]. Applied Energy, 2011, 88(4):1405-1414.

[7] Wang L, Zou H, Su J, et al. An ARIMA-ANN hybrid model for time series forecasting [J]. Systems Research and Behavioral Science, 2013, 30(3):244-259.

[8] Ren C X, Wang C B, Yin C C, et al. The prediction of short-term traffic flow based on the niche genetic algorithm and BP neural network [M]. Berlin:Springer, 2012:775-781.

[9] Liu Z B, Yang Z S, Gao P. Research on the short-term traffic flow prediction method based on BP neural networks [C]//World automation congress. [s. l.]:[s. n.], 2012.

[10] Jia J. Financial time series prediction based on BP neural network [J]. Applied Mechanics and Materials, 2014, 631-632:31-34.

[11] 武志勇, 刘东升. 小波神经网络在液压装置压力测试中的应用 [J]. 电子测量技术, 2010, 33(12):91-94.

[12] Khashei M, Bijari M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting [J]. Applied Soft Computing, 2011, 11(2):2664-2675.

[13] 王 堃, 陈涛涛, 李 雪, 等. 基于 Matlab 的小波神经网络参考作物腾发量预测模型研究 [J]. 沈阳农业大学学报, 2013, 44(4):457-460.

[14] 赵成柏, 毛春梅. 基于 ARIMA 和 BP 神经网络组合模型的我国碳排放强度预测 [J]. 长江流域资源与环境, 2012, 21(6):665-671.

[15] 熊志斌. 基于 ARIMA 与神经网络集成的 GDP 时间序列预测研究 [J]. 数理统计与管理, 2011, 30(2):306-314.